UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

CHEMIMAGE CORPORATION, *Plaintiff,* v. JOHNSON & JOHNSON and ETHICON, INC., *Defendants*.	Civil Action No. 21:24-cv-2646 (JMF)

AFFIDAVIT OF JOSEPH PETER CORRIGAN

I, Joseph Peter Corrigan declare as follows:

1. I submit this affidavit to provide direct testimony in this action.

2. I have personal knowledge of the facts stated below. If called to testify, I could and would testify to the recited facts.

3. I am over eighteen years of age. I am competent to testify regarding the matters set forth in this affidavit.

I. EDUCATION AND EXPERIENCE

4. I am employed at Cambridge Consultants, Ltd. ("CC"). I have worked at CC since September 2017.

5. I am Head of Technology for CC's Intelligent Services and Healthcare business units. In this position, I am responsible for overseeing CC's artificial intelligence, machine learning ("AI/ML") and digital health projects.

6. I graduated from The University of Manchester Institute for Science and Technology with a degree in Mechanical Engineering (Meng).

7. After graduation, I have the following work experience:

   a. 4 years in medical device consulting including statistical modelling and machine control;

   b. 8 years with PlaqueTec Ltd., where I used statistical inference, machine learning and computer vision to successfully design, clinically trial, CE mark and demonstrate feasibility of a cardiovascular biomarker discovery catheter;

   c. 1 year for Cazana using machine learning for image processing and text analysis (winning the BMW Innovation lab and subsequently selected by the UK's financial regulator to settle insurance claims); and

d. 7 years at CC growing the AI/ML in Healthcare capability, including leading our teams to develop AI/ML approaches for detection of tuberculosis in images, medical device optical inspection, endoscopy and continuous glucose monitoring amongst other projects.

8. In total, I have approximately 15 years' experience with AI/ML, and I have worked on over 50 AI/ML projects including ophthalmic vision systems, surgical vision systems, endoscopic vision systems, continuous blood glucose meters, optical coherence tomography systems, ultrasound systems, critical care systems, cardiovascular catheters, cardiovascular imaging systems.

9. CC is a product development and technology consultancy. We employ approximately 800 people, over 90% of which are engineers, technologists, designers, scientists and technical consultants. We typically provide independent advice to clients on complex engineering, science and technology projects, and often serve as an independent technical reviewer.

10. Due to the nature of CC's work, we typically assemble multidisciplinary teams that draw upon our approximately 60 years' experience developing wide range of products across Healthcare, Consumer, Industrial, and Wireless specialties including supporting the development of a wide range of AI/ML enabled healthcare products. CC employs highly competent scientific and engineering staff to help our clients identify and reduce risk in technology for their product development programs. CC is certified to ISO 9001 and ISO 13485.

11. For each client project, we follow appropriate quality management processes and provide frequent client feedback. CC is globally recognized for its AI/ML work, being named a

Forbes top 10 AI consulting firms,[1] and an AIconics Award winner,[2] and has helped clients bring multiple medical devices to market.

II. WORK ON PROJECT ERIE

12. I am familiar with Project Erie, which is a codename Ethicon, Inc. ("Ethicon") used for its review of ChemImage's proposed multispectral imaging system that Ethicon wanted to use for critical structure detection and cancer localization during laparoscopic and robotic medical procedures.

13. Ethicon retained CC to provide independent technical review services and guidance on Project Erie. In this role, CC evaluated data ChemImage presented to Ethicon to attempt to demonstrate achievement of certain project milestones and to inform Ethicon of potential project risks.

14. I first learned about CC's work on Project Erie in about November 2020. I became personally involved in January 2021, as the project manager overseeing CC's AI/ML review of ChemImage's related work through several sub-projects including CC's *Ignore Labels Report* discussed below.

15. The following terms used in this affidavit, and the referenced CC reports, are defined in the context of Project Erie as follows:

a. Area under the [ROC] curve (AUC or AUROC) is a statistical measurement of how well a diagnostic test or model can discriminate between two classes (e.g. cancer vs not cancer). An AUC of 1 is a perfect model, meaning that in all cases

[1] *What Are The 10 Best AI Consulting Firms*, Bernard Marr, Forbes, August 17, 2022; https://www.forbes.com/sites/bernardmarr/2022/08/17/what-are-the-10-best-ai-consulting-firms/.
[2] *AIconics Awards San Francisco 2019: Winners Announced*, Max Smolaks, AI Business, September 24, 2019, https://aibusiness.com/verticals/aiconics-awards-san-francisco-2019-winners-announced#close-modal (see best innovation in Deep Learning).

the model determines the correct result. An AUC of 0.5 means that the probability that the model is correct is 50:50, i.e. equivalent to a random guess.

b. Best case performance was used by the parties in this instance to mean the upper bound on performance metrics demonstrated with a few examples of good performance. In other words, a metric meaning the maximum achievable performance, which is not representative of the average performance or the range of performance. When trying to mitigate a risk in a medical device we need the probability of the occurrence of harm to not exceed an acceptable threshold which means that we want to know that the probability of occurrence of the *worst-case* performance is lower than an acceptable threshold (e.g. there is a < 1 in 10,000 chance of malfunction that could cause harm).

c. Consistent case performance was used by the parties in this instance to mean performance demonstrated with acceptable statistical rationale, usually as a mean (average) plus a stated variance (or standard deviation), or a statistically defined probability of exceeding a threshold. The aim here is to demonstrate that the system produces high quality repeatable results and that the variation in these results is predictable, well understood and unlikely lies between, above or below acceptable thresholds.

d. True Negative (TN) is a correct prediction of a negative condition or attribute.

e. True Positive (TP) is a correct prediction of a positive condition or attribute.

f. False Negative (FN) is an incorrect prediction of a negative condition or attribute.

g. False Positive (FP) is an incorrect prediction of a positive condition or attribute.

h. Ignore label is a type of segmentation label used to instruct the segmentation algorithm not to analyze or classify these regions, essentially excluding them from the segmentation process.

i. Negative predictive value (NPV) is a statistical measure that indicates the proportion of negative results that are true negative results.

j. Positive Predictive Value (PPV) is a statistical measure that indicates the proportion of positive results that are true positive results.

k. Scoring is the process of assessing the performance of a model.

l. Sensitivity is a descriptor of test accuracy that describes the true positive rate This is normally expressed in terms of true positives and true negatives as: (number of true positives/(number of true positives + number of false negatives)), this metric gives an indication of the proportion of true positive pixels that have been correctly classified. For example, a low score would indicate that a high proportion of an artery would not be highlighted to the user, risking misleading the user into thinking it would be safe to perform surgery when it is not, with potentially catastrophic implications. Major vascular injuries are responsible for between approximately 75-80% of laparoscopy-associated fatalities.

m. Specificity is a descriptor of test accuracy that indicates the true negative rate. This is normally expressed in terms of true negative/positives as: (number of true negatives / (number of true negatives + number of false positives)). For example, a low specificity with arteries would indicate that a detection system would likely give false indicators of the presence of vasculature. In combination

with a low sensitivity (as in our example in l.), this could further mislead the user into performing unsafe actions as not only is the artery hidden to the user, but the system has produced a plausible alternative further increasing the surgeon's false confidence.

n. Training data is the set of data used to train a machine learning model.

o. Test data is the set of data used to give an unbiased evaluation of the final model fit on the training data.

III. MILESTONE 1A REPORT

16. I understand that, in October 2020, CC conducted an independent review of ChemImage's Milestone 1A final report.[3]

17. The *Milestone 1A Report* attached as Exhibit A to this affidavit is an accurate copy of an original business record CC made, kept in the regular course of our business, and delivered to Ethicon. Preparation of such a report by CC typically involves consultation with a client to define key questions to be answered, formation of a suitable team, gathering and analysis of information, frequent progress, update and clarification meetings with a client and/or partner company (biweekly, monthly), and delivery of draft and final reports as appropriate.

18. Although I did not directly participate in the *Milestone 1A Report* work, I am familiar with CC's *Milestone 1A Report* content through my role at CC, including our analysis and conclusions.

[3] CC's *Erie Milestone 1A Summary Report*, October 9, 2020 ("*Milestone 1A Report*") DEF_00550247-283 attached as Exhibit A.

19. For example, I understand that, as part of this Milestone 1A independent review work, CC reviewed the "technical validity" and "potential risks related to Milestone 1A," as well as identified "technical risks, gaps, and mitigation" "going into Milestone 1B."[4]

20. During our review, CC found and reported to Ethicon that "there were far fewer successes than expected as well as technical challenges that suggest caution."[5] CC identified depth of detection as "the most challenging requirement to achieve"; specifically, "evidence to support detection up to (and beyond) 5 mm [wa]s limited" and "[d]etection beyond 6 mm lack[ed] compelling evidence."[6] CC also commented on the limited Milestone 1A data of twelve scenes to support depth of detection in the 0-5 mm range.[7]

21. CC concluded that although the Milestone 1A acceptance criteria was "considered satisfied as stated and interpreted by ChemImage…"[8] "there [wa]s consensus that these criteria are not a strong enough basis for product development and that there are still significant residual risks that must be addressed before feasibility can be claimed"

22. From these findings, CC recommended the parties "mitigate the residual technical risk" by implementing an interim milestone before proceeding to the next milestone, Milestone 1B.[9]

23. CC also recognized that while "[b]est case performance was above required specification (ChemImage's interpretation)…consistent case performance (Cambridge Consultants' interpretation) was relatively poor and PPV is unacceptable". The "aim" going

[4] *Milestone 1A Report,* DEF_00550247 at 251.
[5] *Milestone 1A Report*, DEF_00550247 at 249.
[6] *Milestone 1A Report*, DEF_00550247 at 258.
[7] *Milestone 1A Report*, DEF_00550247 at 258.
[8] *Milestone 1A Report*, DEF_00550247 at 249.
[9] *Milestone 1A Report*, DEF_00550247 at 268.

forward should be to "move from the achieving best case performance to consistency of performance across a sufficiently ample spectrum of biological variability."[10]

24. My impression is that the detection and consistency issues CC identified when reviewing ChemImage's Milestone 1A final report data persisted throughout the rest of the project. For example, in our January 13, 2023 review of ChemImage's model performance, although we were limited to black box testing (a weaker form of testing than having full access to the model) we were still able to identify a "sharp drop-off in performance at 2nd degree obscuration" and that "context appears to override spectral content" indicating that the model was likely using contextual clues (adjacent unobscured tissue) to indicate the presence of obscured tissue instead. [11] In this work it appeared that ChemImage was unable to demonstrate that its hyperspectral technology could independently identify obscured critical structures—or do so in a consistent way without the help of adjacent unobscured tissue.

25. The apparent struggle of ChemImage's system to perform with obscured tissue is noteworthy because, based on my understanding, a main goal of ChemImage's hyperspectral imaging technology was to assist in the detection of potentially obscured critical structures.

IV. IGNORE LABELS REPORT

26. I was the project manager for CC's independent review of ChemImage's Final *Milestone 1B: Veins, Arteries, and Bile Ducts (VAB) Report*, dated December 16, 2022

[10] *Milestone 1A Report*, DEF_00550247 at 265, 269.
[11] *Review of Source of Segmentation Information Report,* P4292-R-003 v1.0 (January 13, 2023) ("*Segmentation Report*"), DEF_0053061 at 119 attached as Exhibit B, previously marked as Exhibit 11 to the November 14, 2024, Glover Deposition ("Glover Deposition").

("*ChemImage's 1B VAB Final Report*"). My role was to assemble a CC team to evaluate this ChemImage report and ensure that the quality of output met our and client's expectations.

27. During our review, Ethicon specifically asked CC to assess the impact of ChemImage's use of ignore labels on scoring data presented by ChemImage, including both the theoretical impact on scoring metrics and the actual impact as implemented in *ChemImage's 1B VAB Final Report*.

28. To assess the impact of ChemImage's use of ignore labels, CC analyzed the ChemImage data provided to us, described the mechanisms of the metrics used and used a sample of examples from *ChemImage's 1B VAB Final Report* to assess the strength and validity of claimed scoring metrics and identify any observed risks as appropriate.

29. CC sent Ethicon an initial draft of our *Ignore Labels Report* and conclusions on March 1, 2023.[12] Ethicon provided feedback, and CC added some explanatory wording to clarify certain points, but our *Ignore Labels Report* conclusions did not change in the final report issued to Ethicon on May 19, 2023.[13]

30. The *Ignore Labels Report v0.1w* and the *Ignore Labels Report v1.1* attached as Exhibits C and D are accurate copies of business records CC made, kept in the regular course of our business and delivered to Ethicon. Our preparation of the *Ignore Labels Report* involved similar steps as outlined above with respect to the *Milestone 1A Report*.

---

[12] *Independent Review—Ignore Labels Report,* P4292-R-004 v0.1w (February 13, 2023) ("*Ignore Labels Report v0.1w*"), DEF_00570108 at 2 (slide pagination), attached as Exhibit C, previously marked as Exhibit 6 to the Glover Deposition Glover Deposition.

[13] *Independent Review—Ignore Labels Report,* P4292-R-004 v1.1 (May 17, 2023) ("*Ignore Labels Report v1.1*"), DEF_00804843-913 at 844, attached as Exhibit D, previously marked as Exhibit 8 to the Glover Deposition.

31. Before addressing our conclusions in the *Ignore Labels Report*, I want to briefly describe certain well-known mechanics of AI/ML training. As part of AI/ML dataset creation, training data may need to be cleaned to ensure that algorithms are trained on relevant data and ignore labels may be used when labelling training data to exclude certain irrelevant data from the model training. For example, fingers or gauze may be present in training data but would not appear in a laparoscopic surgery operating environment and that data is therefore irrelevant for the system model to learn. In this example, images of fingers or gauze present in training data may be marked with an ignore label, as those items will not appear in the ultimate operating environment. Accordingly, careful thought must be given to data handling and cleaning procedures, including a system model's use of ignore labels, to ensure that bias is not introduced. These considerations are often referred to in AI/ML as Good Machine Learning Practices.

32. For example, in October 2021 "The U.S. Food and Drug Administration (FDA), Health Canada, and the United Kingdom's Medicines and Healthcare products Regulatory Agency (MHRA) have jointly identified 10 guiding principles that can inform the development of Good Machine Learning Practice (GMLP). [Use of these] guiding principles will help promote safe, effective, and high-quality medical devices that use artificial intelligence and machine learning (AI/ML)." Relevant considerations for CC's *Ignore Labels Report* include GMLP point 4, which states: "**Training Data Sets Are Independent of Test Sets:** Training and test datasets are selected and maintained to be appropriately independent of one another. All potential sources of dependence, including patient, data acquisition, and site factors, are considered and addressed to assure independence." And, GMLP point 8, which states in part: "**Testing Demonstrates Device Performance During Clinically Relevant Conditions**: Statistically sound test plans are

developed and executed to generate clinically relevant device performance information independently of the training data set…"[14]

33. CC's analysis of *ChemImage's 1B VAB Final Report* and related documentation led us to observe that "[t]o train segmentation algorithms," ChemImage labelled regions of images as target and non-target and that ChemImage used "ignore labels" to "provide indications of uncertainty in labelling by masking off particular pixels."[15] "During training, ignore labels [were] down-weighted (to an unknown amount) compared to background (for general ignores) and target (for target ignores)."[16]

34. Furthermore, ChemImage did not use ignore labels solely for training data purposes; but it also applied ignore labels to the test data used to calculate its performance metrics.[17] Specifically, during scoring, ChemImage removed pixels with ignore labels from calculations of performance metrics.[18]

35. The goal of CC's analysis was to understand and communicate to Ethicon the potential impact that the use of ignore labels in *ChemImage's 1B VAB Final Report* data could have on performance metrics because these metrics "provide the high level means to understand the progress in the development of the Erie technology", "are likely to be used as an indicator of user experience and medical utility", and "are also likely to be used … to support certification of the Erie technology as a medical device".[19] In other words, Ethicon wanted us to investigate and understand whether ignore labels impacted performance metrics in a way that the reported metrics

---

[14] *Good Machine Learning Practice for Medical Device Development: Guiding Principles*, ("GMLP") (October, 2021), attached as Exhibit E, at 1-2.
[15] *Ignore Labels Report v1.1*, DEF_00804843 at 844.
[16] *Ignore Labels Report v1.1*, DEF_00804843 at 853.
[17] *Ignore Labels Report v1.1*, DEF_00804843 at 850, 853, 854.
[18] *Ignore Labels Report v1.1*, DEF_00804843 at 854.
[19] *Ignore Labels Report v1.1*, DEF_00804843 at 859.

might no longer show development progress, indicate user experience and medical utility, and/or be used to support medical device certification.

36. To better understand whether ChemImage's use of ignore labels impacted reported performance metrics, CC analyzed a subset of the dataset used in the *ChemImage's 1B VAB Final Report* to investigate their practical effects.[20] I believe CC had sufficient data to reach the conclusions set forth in our report. For example, to identify examples where use of ignore labels would have affected performance metrics, we only needed to identify illustrative examples. Since we found multiple examples where performance metrics could have been affected, we did not need additional data to identify and explain the impact of this risk.

37. Based on the data provided, and our investigation and analysis set forth in the *Ignore Labels Report*, CC concluded:

> "***We can therefore conclude as a direct result of the implementation of ignore labels, there is a significant risk that the metrics are misleading indicators of performance and the report out cannot be evaluated on the basis of these metrics***."[21]

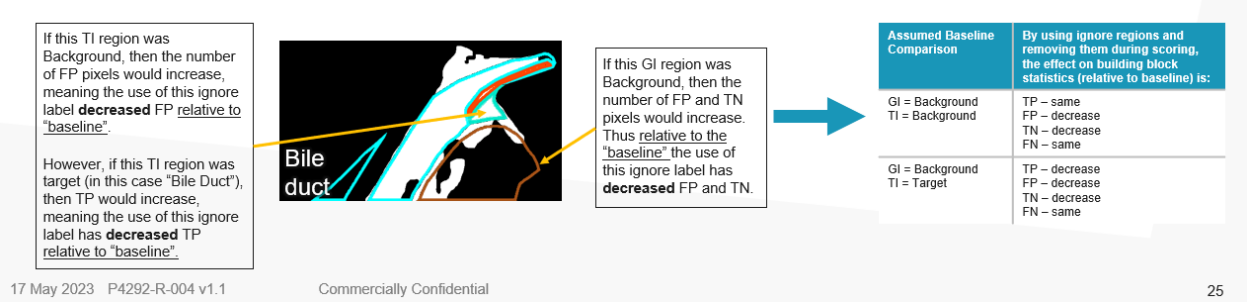
38. In other words, because ignore labels were used to mask regions in the test data used for calculating metrics, and the indicated detection regions overlap with these regions, some portion of the pixels indicating detection or otherwise were not being counted or correctly assigned and therefore the scoring metrics were not correctly indicative of the Project Erie system performance.

39. In our analysis, we provided several examples at pages 27-54 of our *Ignore Labels Report* showing that, as ignore label pixels were excluded from scoring, any removal of pixels or overlap between the detection output of the algorithm and the ignore regions affected the number

[20] *Ignore Labels Report v1.1*, DEF_00804843 at 905.
[21] *Ignore Labels Report v1.1*, DEF_00804843 at 844 (emphasis added).

of pixels included in the components of the summary metric calculations. We could and did therefore conclude that the calculated summary metrics were affected. Because the observed ignore label regions (i.e., the removed and overlapping pixels) were sufficiently large that they could be assessed visually, we were able to make a qualitative assessment of their effect on the metrics without calculating the area of overlap, and the magnitude of the effect did not need to be quantified. Where a pixel count would be required to provide a quantitative assessment, we indicated this with "quantification required".

40. For clarity we created an explanatory image on page 25 of our *Ignore Labels Report* showing how we assessed the qualitive effect on the metrics, which is reproduced below:[22]

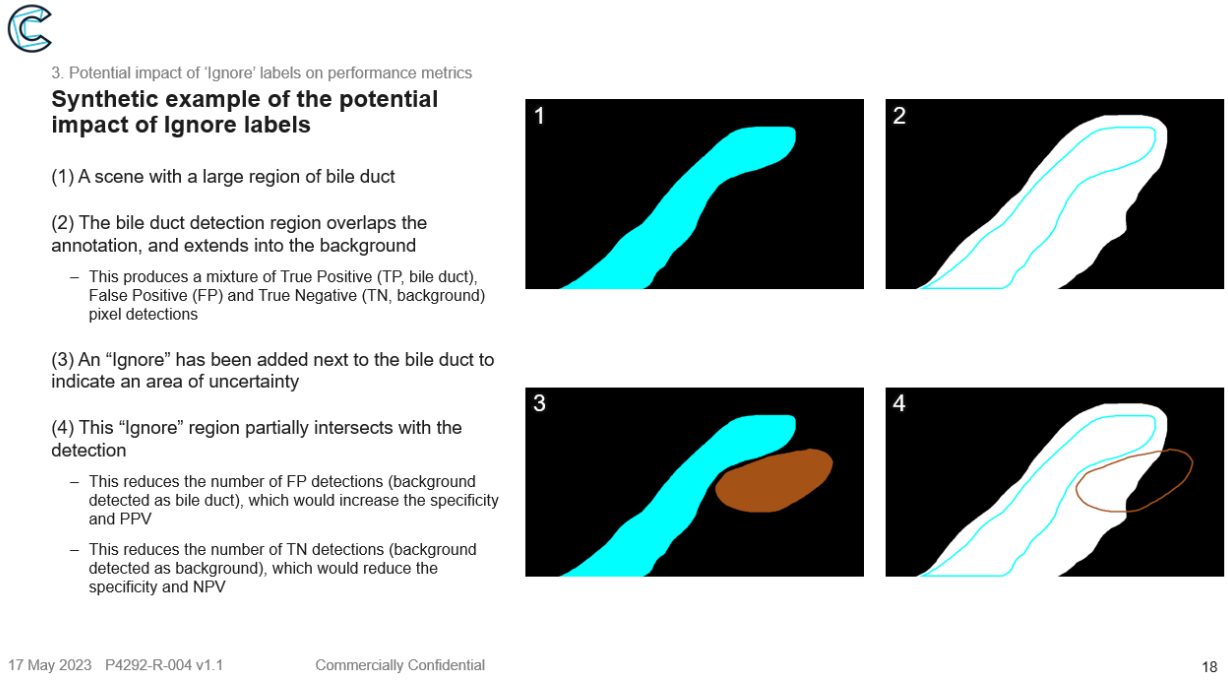


41. CC identified fourteen examples from the supplied ChemImage dataset where excluding pixels within ignore label regions (outlined in brown in the snapshot image above) from scoring calculations would be expected to alter performance metrics.[23] Specifically, CC "identified multiple examples in the data where the use of ignore labels has undermined the validity of multiple metrics by removing pixels of indeterminate status."[24] In other words where these ignore labelled pixels are removed from calculations, "[t]his reduces the [number of] pixels that are

[22] *Ignore Labels Report v0.1w*, DEF_00570108 at 25 (slide pagination).
[23] *Ignore Labels Report v1.1*, DEF_00804843 at 905, 869-898; *see also*, *Ignore Labels Report v0.1w*, DEF_00570108 at 63, 27-56 (slide pagination).
[24] *Ignore Labels Report v1.1*, DEF_00804843 at 844.

counted by the detection building block statistics (TP, FP, TN, FN), which in turn impacts the overall performance metrics"[25].

42. We created a synthetic example to help explain this impact on page 18 of our *Ignore Labels Report*, which is reproduced below:[26]



43. This synthetic example references the calculations for the main metrics, which are listed on page 15 of our *Ignore Labels Report*, demonstrates the impact introduced by ChemImage's use of ignore labels in scoring test data. For example, if we examine the calculation metric for "Specificity = TN/(TN+FP)" we can see that use of an ignore label (as shown and outlined in brown in Figures number 3 and 4 snapshot image above) removes True Negative (TN) pixels (i.e. the pixels were correctly determined by the detection region to be background), from the calculated Specificity metric. To explain this better we can work through an example with

[25] *Ignore Labels Report v1.1*, DEF_00804843 at 858.
[26] *Ignore Labels Report v0.1w*, DEF_00570108 at 18 (slide pagination).

illustrative numbers, say for example we started with 100 TN and 100 False positive (FP) pixels in our image region, the value for Specificity = 100/(100+100) = 0.5. If we applied the ignore label so that it masked off half of the FP pixels (e.g. similar to the white part inside the brown outline in image 4 of our synthetic example above), the calculation becomes Specificity = 100/(100+50) = 0.66 – so in this example the Specificity metric is increased by applying the ignore label and removing the associated pixels from the calculation – making the resulting calculated Specificity metric a misleading indicator.

44. The use of ignore label regions as shown above is not consistent with GMLP.[27] The above use of ignore labels in the ChemImage test scoring data allows potentially clinically representative data to be removed from the test set, which is not representative of clinical use. This means the metrics provided by ChemImage were not representative of clinical use or a true representation of ChemImage's system performance.

45. CC did not conduct a quantitative analysis, which would have been necessary to "determine if the effect of the use of Ignore labels across the overall dataset would act to increase or decrease the top-level performance metrics."[28] Rather, CC "determined that removing pixels within ignore regions from scoring calculations will *always impact* at least one of the performance metrics, regardless of where in the scene the Ignore label is placed,"[29] and "it would not be possible to remove pixels within Ignore regions from scoring without impacting at least one of the potential metrics, regardless of where in the scene the Ignore label is placed."[30] In other words, as all of the

[27] GMLP, at 2.
[28] *Ignore Labels Report v1.1*, DEF_00804843 at 848.
[29] *Ignore Labels Report v1.1*, DEF_00804843at 848 (emphasis added).
[30] *Ignore Labels Report v1.1*, DEF_00804843 at 905.

metrics are derived from pixel counts, changing the total pixel count by masking them with ChemImage's use of ignore labels will always affect at least one of the metrics.

46. "The true state of tissue within 'Ignore' and 'Ignore-target' regions [was] unknown, so it [was] not possible to evaluate the impact of detections (or lack of) in [those] regions[.]"[31] Because CC was not aware of ChemImage's process for applying ignore labels, CC had, and raised to Ethicon, concerns about what critical structure ChemImage's technology detected (or missed).

47. Specifically, CC concluded and reported to Ethicon that ChemImage's use of ignore labels on scoring data "introduce[d] uncertainty around the ability of the performance metrics to describe the quality of segmentation results" and "there [was] a significant risk that the metrics [we]re a misleading indicator of performance."[32]

48. CC's work on Project Erie ended shortly after our submission of the final *Ignore Label Report*. Later, and some time after submission of our final *Ignore Label Report*, I learned that Ethicon had terminated Project Erie with ChemImage.

V. MISCELLANEOUS

49. I was not involved with or aware of Ethicon's decision to terminate the Agreement. I am not aware of anyone else at CC's being involved in that decision.

50. As far as I know, CC was not involved with or aware of any decisions made by the Project Erie Data Review Board.

[31] *Ignore Labels Report v1.1*, DEF_00804843 at 905.
[32] *Ignore Labels Report v1.1*, DEF_00804843 at 905.

VI. CONCLUSION

51. CC was an independent technical consultant for Project Erie. CC's conclusions as set forth in its *Milestone 1A Report* and *Ignore Label Report* were not influenced by Ethicon, Johnson & Johnson or any other person or entity.

52. CC stands by the conclusions in its *Milestone 1A Report* and *Ignore Labels Report*.

VII. RESERVATION OF RIGHTS

53. I reserve the right to amend or supplement this affidavit as this case proceeds and any additional information may come to my attention.

54. I hereby declare that all statements made in this affidavit of my own knowledge are true and that all statements made on information and belief are believed by me to be true. I further declare that the statements I make in this affidavit are made with the knowledge that willfully false statements made to United States Courts are punishable by fine or imprisonment, or both, under Section 1001 of the Title 18 of the United States Code.

I declare under penalty of perjury that the foregoing statements are true and correct.

Dated: 30/1/2025

Joseph Peter Corrigan

Sworn by the said Joseph Peter Corrigan, identified to me by his United Kingdom passport number 132678706 at 20 Station Road, Cambridge CB1 2JD, United Kingdom on 30 January 2025

before me

DUNCAN EDWARD WALKER

DUNCAN WALKER NOTARY PUBLIC

Duncan Walker
Notary Public
20 Station Road
Cambridge CB1 2JD
England
+44 1223 753676
Protocol number: 212184.0001